

UNCLASSIFIED

AD 407 039

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

CATALOGED BY DDC

AS AD No. \_\_\_\_\_

407039

407 039

THE DEVELOPMENT OF STATISTICAL METHODS  
FOR QUALITY CONTROL AND SURVEILLANCE TESTING

SUBJECTIVE TESTING AND QUALITY EVALUATION

Research supported by the Army, Navy and Air Force under  
Office of Naval Research Contract Number NONR-988(08),  
Task Order NR 042-004 with the Florida State University.  
Reproduction in whole or in part is permitted for any  
purpose of the United States Government.

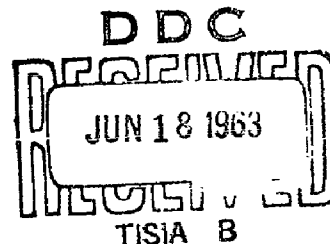
Technical Report No. 7

SEQUENTIAL RANK TESTS

by

RALPH A. BRADLEY

April, 1963  
Department of Statistics  
Florida State University  
Tallahassee, Florida



NO.OTS

# FIFTH INTERNATIONAL BIOMETRIC CONFERENCE

CAMBRIDGE, 1963

## Sequential Rank Tests<sup>1</sup>

by Ralph A. Bradley

The Florida State University

### 1. Introduction and Summary.

Wilcoxon [1945] introduced the two-sample, rank-sum test for testing the difference in locations for two populations. Consider  $m$  observations,  $X_1, \dots, X_m$  from an  $X$ -population and  $n$  observations,  $Y_1, \dots, Y_n$  from a  $Y$ -population, all observations being independent. If  $F(u) = P(X \leq u)$  is the cumulative distribution function (c.d.f) of the  $X$ -population and  $G(u) = P(Y \leq u)$  is the c.d.f. of the  $Y$ -population, the hypothesis tested is  $H_0: F(u) \equiv G(u)$  versus the alternative (one-sided for illustration),  $H_a: F(u-a) \equiv G(u)$ ,  $a > 0$ . The procedure is to rank all observations in joint array yielding sets of ranks  $r_1, \dots, r_m$  and  $s_1, \dots, s_n$  corresponding to the observations and to obtain the sum of ranks, say,  $\Sigma s$  for the  $Y$ -sample. Let  $C(m, n, \alpha)$  be the smallest positive integer for which  $P(\Sigma s \geq C | H_0) \leq \alpha$ ; note that, under  $H_0$ , all configurations of  $X$ 's and  $Y$ 's in the joint array are equally likely. When  $\Sigma s \geq C$ ,  $H_0$  is rejected in favor of  $H_a$  and the significance level is  $\alpha$ . Much has been written about the rank-sum test and various tables of values of  $C$  have been prepared. An excellent bibliography on nonparametric tests has been prepared by Savage [1962].

Lehmann [1953] has proposed an alternative to  $H_0$  different from that above using  $G(u) \equiv F^k(u)$ . Thus, given this model, we write  $H_0: k = 1$  and  $H_a: k > 1$ . Values of  $k > 1$  lead to a change in location of the  $Y$ -population but also to changes in shape of  $G(u)$  relative to  $F(u)$ . Savage [1956] further discussed the implications of the Lehmann model. The basic advantage of the model

---

<sup>1</sup>Research supported by the Army, Navy, and Air Force under an Office of Naval Research Contract. The assistance of a National Science Foundation Grant to the Florida State University Computing Center is acknowledged. Reproduction in whole or in part is permitted for any purpose of the United States Government.

is that it permits relatively easy calculation for the probability of each particular configuration of X's and Y's in the joint array under the alternative hypothesis. Any such configuration may be defined by the ranks assigned to the Y-sample and the result is that

$$P(s_1, \dots, s_n | m, n, k) = \frac{k^n}{\binom{m+n}{n}} \prod_{j=1}^n \frac{\Gamma(s_j + jk - j) \Gamma(s_j + 1)}{\Gamma(s_j + 1 + jk - j) \Gamma(s_j)}, \quad (1)$$

interpretable given that  $s_{n+1} = m + n + 1$ ,  $1 \leq s_1 < \dots < s_n \leq m + n$ . It is an additional property of the model that

$$p = P(X \leq Y) = k/(k + 1). \quad (2)$$

The Lehmann model permits the development of two-sample, sequential rank tests. Initial work on this problem was done by Wilcoxon, Rhodes and Bradley [1963] and is summarized in the next section. The standard sequential probability-ratio method of Wald [1947] was used for pairs of samples of sizes  $m$  and  $n$  taken sequentially, the ranking effected within each group of two samples. In this paper we examine certain extensions of that sequential procedure. In addition, the Lehmann model may not be the one desired in practice and departures from the model are considered. At the time of preparation of this paper, much of the research was still in progress and results given here are in some cases preliminary and conclusions somewhat tentative.

## 2. Two Sequential Two-Sample Grouped Rank Tests.

Let a group of observations consist of  $m$  X-observations and  $n$  Y-observations as discussed above. In Wald sequential analysis, we shall take the group as the requisite unit to be taken sequentially. For each group the probability ratio is required and obtainable from (1). This ratio for the  $\gamma$ -th group is

$$r_{\gamma}(m, n, k_1, 1) = k_1^n \frac{(m+n)!}{\Gamma(s_{1,\gamma})} \prod_{j=1}^n \frac{\Gamma(s_{j,\gamma} + jk_1 - j)}{\Gamma(s_{j+1,\gamma} + jk_1 - j)} \quad (3)$$

where  $s_{j,\gamma}$  is the rank of the  $j$ -th Y in the  $\gamma$ -th group and (1) is used in the denominator with  $k = 1$  under the null hypothesis  $H_0$  and in the numerator with  $k = k_1 > 1$  under  $H_a$ . Note that this probability ratio is dependent upon the configuration of X's and Y's in the joint array for the  $\gamma$ -th group and we have designated the sequential test based on it as the configural rank test. If one

is at the  $t$ -th stage of a sequential configural rank test, the test statistic is

$$r_Y(m, n, k_1, 1) = p_{1t}/p_{ot} \quad (4)$$

in the notation of Wald. Suppose that  $\alpha$  has been specified as the probability of a Type I error and  $\beta$  as that for a Type II error. Then the sequential decision procedure is to

- (i) Terminate the test with the rejection of  $H_0$  (acceptance of  $H_a$ ) if  $p_{1t}/p_{ot} \geq A = (1 - \beta)/\alpha$ ,
- (ii) Terminate the test with the acceptance of  $H_0$  if  $p_{1t}/p_{ot} \leq B = \beta/(1 - \alpha)$ , or

- (iii) Consider another group of observations if  $A < p_{1t}/p_{ot} < B$ .

Wilcoxon has suggested a simple algorithm for the computation of (1) and then (3) and (4) which is explained and illustrated by Wilcoxon, Rhodes and Bradley [1963].

A second two-sample grouped sequential rank test is based on the within-group sums of ranks for the  $Y$ -sample. Let  $S_Y$  be that rank sum for the  $Y$ -th group,  $S_Y = \sum_{j=1}^n s_{j,Y}$ . Now, for given  $k$ ,

$$P\left(\sum_{j=1}^n s_{j,Y} = S_Y \mid m, n, k\right) = \sum_{\substack{1 \leq s_{1,Y} < \dots < s_{n,Y} \leq m+n \\ \sum_{j=1}^n s_{j,Y} = S_Y}} P(s_{1,Y}, \dots, s_{n,Y} \mid m, n, k) \quad (5)$$

where the argument of the sum comes from (1). The probability ratio statistic for the  $Y$ -th group for the sequential rank-sum test is

$$R_Y(m, n, k_1, 1) = P\left(\sum_{j=1}^n s_{j,Y} = S_Y \mid m, n, k_1\right) / P\left(\sum_{j=1}^n s_{j,Y} = S_Y \mid m, n, 1\right). \quad (6)$$

At the  $t$ -th stage of the sequential rank-sum test, the appropriate probability-ratio statistic is

$$p_{1t}/p_{ot} = \prod_{Y=1}^t R_Y(m, n, k_1, 1) \quad (7)$$

and the decisions noted above for  $p_{1t}/p_{ot}$  apply to  $P_{1t}/P_{ot}$  also. The sequential

rank-sum test is easier to apply than the configural rank test when tables are available. Both tests are facilitated through the use of logarithms on both the probability ratios and the sequential bounds. Tables are given in the reference for  $m = n = 1(1)9$ ,  $k_1 = 1.5, 2.33, 4$  and  $9$  showing  $S_Y; T_Y$ , the corresponding rank-sum for the  $Y$ -sample;  $P_Y$  in (5); and  $\ln R_Y$  from (6). The parameter  $p$  of (2) corresponding to values of  $k_1$  has values  $.6, .7, .8$  and  $.9$ . Given  $S_Y$  for the  $y$ -th group, it is now possible to go directly to  $\ln R_Y$ , to sum such values for  $t$  groups, and to compare the sum  $\left( \ln P_{1t}/P_{0t} \text{ from (1)} \right)$  with  $\ln A$  and  $\ln B$ .

The Lehmann model will be strange to most users of the method and interpretation is necessary for sensible choice of  $k_1$ . Some insight comes from the corresponding values of  $p$  and additional help results from considering  $\mu_y$ , the change in location in terms of standard deviations, for the situation where  $F(u)$  is a normal c.d.f. Values of  $\mu_y$  for values of  $k_1$  chosen are  $.232, .658, 1.029$  and  $1.485$  respectively. The standard deviation of the  $Y$ -population in this normal case decreases as  $k$  increases and values are  $.701$  when  $k = 4$  and  $.598$  when  $k = 9$  as fractions of the standard deviation for the  $X$ -population.

Properties of these two sequential rank tests follow from results of Wald. It follows that the processes reach a decision with probability one, Average Sample Numbers (A.S.N.'s) and Operating Characteristic Functions (O.C.-functions) have been evaluated and tabulated. Selected results will be given in tables below. It appears that the rank-sum test is almost as good as the configural rank test and is easier to use.

### 3. Modified Sequential Rank Tests.

It appears intuitively that better rank tests might be obtained if complete reranking of the totality of  $X$ - and  $Y$ -observations were effected at each stage of a sequential process. Such a procedure has considerable theoretical interest although practical considerations are likely to dictate within-group ranking in most applications. Merchant [1962] working with Wilcoxon and Bradley considered this problem.

Suppose that  $X$ - and  $Y$ -observations are taken in pairs corresponding to the situation with  $m = n = 1$  above and with no group or pair effect present. Then, at the  $t$ -th stage of such a process, observations  $X_1, \dots, X_t$  and  $Y_1, \dots, Y_t$

are ordered in joint array. A modified configural rank test would be based on the statistic

$$p_{lt}/p_{ot} = r(t, t, k_1, 1) = k_1^t \frac{(2t)!}{\Gamma(s_1)} \prod_{j=1}^t \frac{\Gamma(s_j + jk_1 - j)}{\Gamma(s_j + 1 + jk_1 - j)} \quad (8)$$

from (3) and a modified rank-sum test would be based on

$$p_{lt}/p_{ot} = R(t, t, k_1, 1) \quad (9)$$

from (7) and (6). Difficulties in theory now enter since successive values of  $\ln(p_{lt}/p_{ot})$  or of  $\ln(R)$  may be regarded no longer as sums of independent random variables and major assumptions for Wald's sequential analysis do not hold. Additional difficulty in applications may occur with the modified rank-sum test since  $t$  may exceed the maximum tabular value of nine and then values of  $\ln R$  will not be easily available. The Wilcoxon algorithm will assist in the use of the modified configural rank test.

It was decided to proceed with the modified sequential rank tests as though the Wald bounds A and B were still appropriate for  $p_{lt}/p_{ot}$  or  $R$ . Monte Carlo studies reported in part below suggest that this is appropriate. Berk [1962] has also worked on these modified sequential rank tests at Harvard University and has reported that he has shown that the ranks  $s_1, \dots, s_t$  at the  $t$ -th stage are sufficient for the first  $t$  rankings. This work and that by Hall [1962] is enough to justify continued use of the bounds A and B and the statistics of (8) and (9).

#### 4. Monte Carlo Results.

Since Wald formulas for A.S.N.- and O.C.-functions are not applicable for the modified sequential rank tests, Merchant proceeded with Monte Carlo studies on the IBM-709 computer. These studies were done only for the modified configural rank test and the method was as follows.

An odd integral value of the parameter  $k$  was chosen and  $(k + 1)$  random standard normal deviates were generated through use of a subroutine that produces these in pairs. For each such set, the first deviate was taken as an X-observation and the largest of the  $k$  remaining deviates was taken as the Y-observation. In this way the Lehmann model was satisfied but for the special case with  $F(u)$ , a



standard normal c.d.f. As each pair of X- and Y-observations were generated, the totality of X- and Y-observations were reranked, the logarithm of the statistic of (8) was computed, and the value so obtained was compared with  $\ln A$  and  $\ln B$ . For this study, Merchant took  $\alpha = \beta = .05$  and hence  $\ln A = 2.944$  and  $\ln B = -2.944$ . The modified configural rank test was simulated with each experiment carried to a decision with  $H_0: k = 1$  and for values of  $k_1$  in  $H_a: k = k_1 > 1$  with  $k_1 = 1.5, 2.33, 4$  and  $9$  for true values of  $k$  of  $1, 3, 5, 7$  and  $9$ . This procedure permitted at least crude graduation of both the A.S.N.-functions and the O.C.-functions. Values of the A.S.N.-function were computed as the average number of trials required for a decision as 500 simulated experiments were conducted for each true value of  $k$  for each sequential design. Results for the values of  $k$  indicated are shown in Column 2 of Table 1 for the sequential design with  $k_1 = 4$  to indicate the nature of results obtained and all A.S.N.-values are in terms of numbers of observations taken from each population. In the same way values of the O.C.-function are in Column 2 of Table 2; these values are simply the proportions of sets of 500 experiments that led to the acceptance of  $H_0$ . Note that the empirically obtained value of  $\alpha$  is .034, less than the nominal value of .05; in general it appears that true values of  $\alpha$  and  $\beta$  are less than the nominal ones of the sequential design. It is also observed that the Wald method appears to be appropriate on the basis of the Monte Carlo method.

Wilcoxon, Rhodes and Bradley gave A.S.N.- and O.C.-values using Wald's formulas. Examples are shown for the grouped rank tests with  $m = n = 4$  for configural and rank-sum tests in Columns 3 and 4 respectively of Tables 1 and 2. Values shown are sparse but suggest a discrepancy between these values and those obtained by Merchant for the supposed better method. It turns out that the Wald formulas underestimate the A.S.N.-values and Monte Carlo results based on sets of 500 experiments for the rank-sum test are given in Columns 5. The comparison of the modified tests and the grouped tests are confounded by the fact that results in Columns 2 are for the modified configural rank test and in Columns 5 for the grouped rank-sum test. We do, however, believe that these comparisons are indicative of the theoretical advantage of the modified tests.

Since it is often thought desirable to consider a model wherein X- and Y-population differ only in locations, sampling studies were also made with

TABLE 1

## VALUES OF THE A.S.N.-FUNCTION

$$\alpha = \beta = .05 \quad H_1: k = 1, H_a: k = k_1 = 4$$

True Value of k	Monte Carlo Mod. Config. Rank Test Leh. Alt.	Wald Form. Gpd. Config. Rank Test m = n = 4 Leh. Alt.	Wald Form. Gpd. Rank-Sum Test m = n = 4 Leh. Alt.	Monte Carlo Gpd. Rank-Sum Test m = n = 4 Leh. Alt.	Monte Carlo Gpd. Rank-Sum Test m = n = 4 Normal Alt.	Monte Carlo Truncation 5 Gps. Rank-Sum Test m = n = 4 Leh. Alt.
0	-----	2.8	2.8	-----	-----	-----
1	8.78	7.8	9.0	13.90**	13.90**	12.45**
2	-----	13.3*	-----	25.26	25.66	15.88
3	13.09	-----	-----	19.90	23.93	14.86
4	-----	9.3	10.1	14.41	16.74	13.03
5	8.64	-----	-----	12.58	14.22	11.99
6	-----	-----	-----	-----	-----	-----
7	6.98	-----	-----	10.18	12.46	10.10
8	-----	-----	-----	-----	-----	-----
9	6.49	-----	-----	9.59	11.11	9.54
$\infty$	-----	4.3	4.5	-----	-----	-----

\* Actual true value of k is 1.96.

\*\* Based on 1000 experiments -- remainder of tables based on 500 experiments.

TABLE 2

## VALUES OF THE O.C.-FUNCTION

$$\alpha = \beta = .05 \quad H_0: k = 1, H_a: k = k_1 = 4$$

True Value of k	Monte Carlo Mod. Config. Rank Test Leh. Alt.	Wald Form. Gpd. Config. Rank Test m = n = 4 Leh. Alt.	Wald Form. Gpd. Rank-Sum Test m = n = 4 Leh. Alt.	Monte Carlo Gpd. Rank-Sum Test m = n = 4 Leh. Alt.	Monte Carlo Gpd. Rank-Sum Test m = n = 4 Normal Alt.	Monte Carlo Truncation 5 Gps. Rank-Sum Test m = n = 4 Leh. Alt.
0	----	1.00	1.00	----	----	----
1	.566	.95	.95	.979***	.979***	.950***
2	----	.50*	.50**	.484	.543	.500
3	.102	----	----	.106	.154	.180
4	----	.05	.05	.026	.060	.046
5	.000	----	----	.002	.013	.008
6	----	----	----	----	----	----
7	.000	----	----	.000	.006	.000
8	----	----	----	----	----	----
9	----	----	----	.002	.000	.002
$\infty$	----	.00	.00	----	----	----

\* Actual true value of k is 1.96. \*\* Actual true value of k is 2.04.

\*\*\* Based on 1000 experiments--remainder of table based on 500 experiments.

$F(u)$ , the standard normal c.d.f. and  $G(u)$ , the c.d.f. for a normal population with unit variance but mean at  $\mu_y$ . The values taken for  $\mu_y$  were those for the mean if  $G(u) \equiv F^k(u)$ . Thus in the example of Columns 6 of Tables 1 and 2  $\mu_y$  was taken to be 0, .564, .846, 1.029, 1.163, 1.352 and 1.485 corresponding to values of  $k$  of 1, 2, 3, 4, 5, 7, and 9 respectively. It is to be noted that the A.S.N.-numbers are somewhat higher for the translation experiments and that the O.C.-numbers are also higher. In particular for data fitting the Lehmann model, the observed  $\beta = .026$ , less than the nominal .05 while for the data fitting the translation model the observed  $\beta = .060$ . For practical purposes it is not thought that the method is too dependent on applications meeting the Lehmann model.

In order to obtain as much information as possible from the Monte Carlo studies, truncation of the process was also considered. The final columns of Tables 1 and 2 show results obtained with a forced decision after five groups,  $m = n = 4$ , for the grouped rank-sum test. For those experiments not already terminated after five groups,  $H_0$  was accepted when the logarithm of the probability ratio in (6) was negative and  $H_a$  was accepted when it was positive; this appeared to be an acceptable rule due to the symmetry present with  $\alpha = \beta = .05$ . From Table 1 it is seen that savings resulted for middle values of  $k$  particularly (compare Columns 5 and 7) and from Table 2 note that neither  $\alpha$  nor  $\beta$  (.05 and .046 respectively) were seriously inflated.

The results of Tables 1 and 2 have been selected to show effects that we believe to be indicative for a sequential system with realistic values of  $\alpha$ ,  $\beta$ ,  $k_1$  and  $m$ ,  $n$ . Other Monte Carlo results have been obtained but investigations are continuing. It is expected that complete results will be reported at a later date.

## 5. Remarks

It is perhaps not appropriate to make many additional comments in this paper. The grouped sequential rank tests will be available in the cited reference well before the Fifth International Biometric Conference takes place. Results obtained by Merchant are in preparation for publication. Monte Carlo studies are still in progress with this work largely being done by Donald C. Martin working with Wilcoxon and the present author. We conclude simply by

noting that illustrative examples of the grouped sequential rank tests are given by Wilcoxon, Rhodes and Bradley [1963] and we believe that these examples are typical of applications of these methods that may usefully be made.

#### 6. Acknowledgments.

The author is very clearly much indebted to his co-workers on various phases of this research. Dr. Frank Wilcoxon has given major leadership throughout this work and it is a great privilege to be closely associated with him at the Florida State University. Mr. L. J. Rhodes and Miss Sarla D. Merchant have assisted greatly in parts of this study related to their theses for the Master of Science degree. Mr. Donald C. Martin of the University Computing Center has advised on computer programs and, indeed, developed the programs for many of the Monte Carlo studies.

#### REFERENCES

- Berk, Robert [1962]. Private communication. Department of Statistics, Harvard University, Cambridge, Mass.
- Hall, W. J. [1962]. The relationship between sufficiency and invariance with applications in sequential analysis. Part I: Exposition and sequential applications. Unpublished manuscript.
- Lehmann, E. L. [1953]. The power of rank tests. Ann. Math. Stat. 24, 23-43.
- Merchant, Sarla D. [1962]. A Study of Two Sequential Rank Tests. M. S. Thesis, Library, Florida State University, Tallahassee, Florida.
- Savage, I. Richard [1956]. Contributions to the theory of rank order statistics - The two-sample case. Ann. Math. Stat. 27, 590-615.
- Savage, I. Richard [1962]. Bibliography of Nonparametric Statistics, Harvard University Press, Cambridge, Mass.
- Wald, Abraham [1947]. Sequential Analysis. John Wiley and Sons, Inc., New York.
- Wilcoxon, Frank [1945]. Individual comparisons by ranking methods. Biometrics 1, 80-3.
- Wilcoxon, Frank, Rhodes, L. J. and Bradley, R. A. [1963]. Two sequential two-sample grouped rank tests with applications to screening experiments. Biometrics 19, (in press), March Issue.